View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Digital Library of UIN Sunan Ampel

CONTENT VALIDITY AND ITEM ANALISYS OF SEMESTER II ENGLISH FINAL TEST FOR TENTH GRADE STUDENTS OF SMAN 3 SIDOARJO

THESIS

Submitted in Partial Fulfillment of the Requirements for the Degree of Sarjana Pendidikan Islam (S.Pd.I) in Teaching English

PERPUSTAKAAN
IAIN SUNAN AMPEL SURABAYA
No. KLAS: K T-2010 036 PBI
No REG : T-2010/PBI/036
ASAL BUKU :
TANGGAL :

Oleh :

MILLATUL ISLAMIYAH
D05206055

ENGLISH EDUCATION DEPARTMENT
FACULTY OF TARBIYAH
STATE INSTITUTE FOR ISLAMIC STUDIES SUNAN AMPEL
SURABAYA
2010

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

APPROVAL SHEET

This thesis by :

Name :Millatul Islamiyah

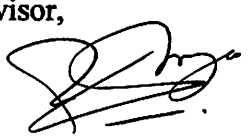
NIM : D05206055

Title : **CONTENT VALIDITY AND ITEM ANALYSIS ON SEMESTER II ENGLISH FINAL TEST FOR TENTH GRADE STUDENTS OF SMAN 3 SIDOARJO**

Has been approved by the advisor and could be proposed to fulfill the requirement for the Graduate Degree of Sarjana Pendidikan in English Department of Tarbiyah Faculty

Surabaya, August 3rd, 2010

Advisor,



Dra. Irma Soraya, M. Pd

NIP. 196709301993032004

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

EXAMINERS APPROVAL SHEET

Thesis entitled:
Content Validity and Item Analysis of Semester II English Final Test for Tenth Grade Students of SMAN 3 Sidoarjo has been accepted and approved by the broad of examiners of English Department of Tarbiyah Faculty State Institute for Islamic Studies Sunan Ampel Surabaya.

Surabaya, August 20th, 2010

Dean of Tarbiyah Faculty,

Dr. H. Nur Hamim, M. Ag
NIP. 196203121991031002

The Board of Examiners
Advisor/Chair Person,

Dra. Irma Soraya, M. Pd
NIP. 196709301993032004

Secretary,

Siti Asmiyah, S. Pd
NIP. 197704142006042003

Examiner I,

Dr. Mohammad Salik, M. Ag
NIP. 196712121934031002

Examiner II,

Masdar Hilmy, M. A, Ph. D
NIP. 197103021996031002

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini adalah:

Nama	: Millatul Islamiyah
Nim	: D05206055
Semester	: VIII (Delapan)
Jurusan	: Pendidikan Bahasa Inggris (PBI)
Fakultas	: Tarbiyah
Alamat	: Pramian Labuhan Sreseh Sampang Madura

Dengan ini menyatakan dengan sebenarnya bahwa skripsi yang berjudul " **Content Validity and Item Analysis of Semester II English Final Test for Tenth Grade Students of SMAN 3 Sidoarjo**", adalah asli dan bukan plagiat, baik sebagian maupun keseluruhannya.

Demikian pernyataan ini sesuai dengan sebenar-benarnya, apabila pernyataan ini tidak sesuai dengan fakta yang ada, maka saya siap dimintai pertanggung jawaban sebagaimana peraturan perundang-undangan yang berlaku.

Surabaya, 3 Agustus 2010
Pembuat pernyataan

Millatul Islamiyah
NIM D05206055

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ABSTRACT

Content Validity and Item Analysis of Semester II English Final Test for Tenth Grade Students of SMAN 3 Sidoarjo

Name : Millatul Islamiyah

NIM : D05206055

Advisor : Dra. Irma Soraya, M. Pd

Key Words: **Content Validity, Item analysis, Index of Difficulty, Index of Discrimination, the Effectiveness of Distractors.**

Testing is one kind of evaluation. As evaluation, testing is very needed to be applied in teaching to know the progress of the students. Without testing, the result of education will be foolish. In order to perform efficient and correctly, testing must be good designed. Nowadays, there are many teachers do not carefully writing a test. They ignore of the criteria of good test, they are content validity, reliability, index of difficulty, index of discrimination, and the effectiveness of distrators. This study is aimed to know what is Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo like in term of content validity, index of difficulty, index of discrimination, and the effectiveness of distractors .

The design used in this study is a descriptive research because it describes the quality of Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo. It also used quantitative approach since it used numerical calculation to compute the data. The object of this study is Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo and only focus on multiple choice items, while the samples are X1, X2, and X3 class which are taken by random sampling.

The result of this study reported that Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo has good content validity. It also reported that the index of difficulty of Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo used for X1 and X2 are acceptable, but they are recognized easy test for X3. Besides, the index of discrimination of this test is for X1, satisfactory for X2, and malfunction for X3. Moreover, the Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo has good distractors for X1, X2, and X3.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

TABLE CONTENT

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

CHAPTER III: RESEARCH METHODOLOGY

CHAPTER IV : DATA ANALYSIS

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

CHAPTER 1

INTRODUCTION

Background of study in this thesis will be settled in this chapter. It will clarify the reason why the writer attempted to analysis the study. Also in this chapter will explain the significant of the study to inform the advantages of the study. Furthermore, it will be followed by other related subtitle to map clearly the scope limitation of the problem, purpose, and the definition of key terms used in this study to avoid misunderstanding and over generalization of the study either.

A. BACKGROUND

English is not only as foreign language but also as international language. It means that English is used for communication among country in the world.

Thus, learning English is really needed moreover since it has two significant.[1] *Firstly*, English is global language which is used in leisure and business either, such as ease of boarder network of relationship and travel to any English speaking country and others without the need of have a translator as well. Furthermore, mastering English language open more job opportunities. Nowadays in the competitive job market it is necessary to speak English. *Secondly*, mastering English language is also needed to access information and knowledge. Today's media such as the Internet, television and the press give you to almost

[1] Troy Landis, *'Motivation and Mastery' Effective Teaching and Learning Method*, Unpublished paper, (Surabaya, 2009), Page 1

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

unlimited access to knowledge and information. In conclusion, having excellent English language ability can raise better chances of getting jobs that pays more and boarder your knowledge and networking.

Considering the importance of English as well as it's a lot of significant, our government through The Department National Education has determined English as one of compulsory content of curriculum taught at junior and senior high school.

> "Kurikulum pendidikan dasar dan menengah wajib memuat: a. Pendidikan agama, b. pendidikan kewarganegaraan, c. bahasa, d. matematika, e. ilmu pengetahuan alam, f. ilmu pengetahuan sosial, g. seni dan budaya, h. pendidikan jasmani dan olahraga, i. keterampilan/kejuruan, j. muatan local".[2]

Then, in regulation attachment of ministry of national education explained that language curriculum contains English language.[3]

English learning aimed in junior high school is oriented to reach functional level. It means that the students should be able to communicate oral and written in their daily life activity. While, English learning in senior high school is expected to reach informational level, because they have been prepared to continue their study in university.[4] Nowadays English language is taught in elementary school even it is introduced earlier to children in the kindergarten. It is aiming to teach English language earlier in order to facilitate children to learn

---

[2] Peraturan Republik Indonesia Tentang Sistem Pendidikan Nasional No. 20 tahun 2003. Chapter X, section 37, verse 1.

[3] Lampiran Peraturan Menteri Pendidikan Tentang Sistem Pendidikan Nasional No. 20 tahun 2003. Page 8.

[4] Depdiknas, *Standard Kompetensi Mata Pelajaran Bahasa Inggris SMP dan Madrasah Tsanawiyah*, (Depdiknas: Jakarta, 2004) page 3

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

easier in the higher level of education. This phenomena show us several efforts conducted by government to provide generation with skillful.

In addition, to measure the success in teaching learning English, evaluation is needed evaluation is very important means in teaching learning process. It is used to know whether the teaching learning activities is success or not. The result conducted from evaluation provides wide information to the teachers to manage their classroom activities, as Dickins and Germaine say:[5]

> "....it can provide a wealth of information to use for the future direction of classroom practice, for the planning of course, and for the management of learning task and students,"

Furthermore, evaluation and teaching is cannot be separated,[6] all teaching process should be followed by evaluation indeed. Without evaluation, it seems impossible to measure as well as report students progress objectively.

Evaluation as Arikunto has said is the process of evaluating teaching learning process.[7] There are several types of evaluation. One of them is test. Test is a series of question of measuring skill, knowledge, intelligences, and capacities of individual or group.[8] Nevertheless it is common that sometimes evaluation considered has same meaning as testing, and that while students are being tested evaluation is taking place. However testing is only one component in the evaluation process.

---

[5] Pauline Rea-Dickins& Kevin Germaine, *Evaluation*,(New York, Oxford University Press, 2008), Page 3

[6] J. B Heaton, *Writing English Language Test*, (New York: Longman Group, 1988). Page 5

[7] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 3

[8] Ibid, Page 29

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Thus, as one tool of evaluation test is needed to be employed in teaching activities. Moreover it has lot of benefits in order to support the success of teaching learning process, such as:

1. To measure language proficiency
2. To diagnose student's strengths and weakness, to identify what they know and what they do not know.
3. To discover how successful student have been in achieving the objectives a course of study.
4. To assist placement of student by identifying the stage or part of a teaching program most appropriate to their ability[9]

Nevertheless, writing tests is not easy as it looks. There are many things of testing techniques should be considerate before going to start writing it in order the test we made is acceptable. There are several conditions of the test where only they are existed; the test is able to call a good test and adequate, such as the test must fulfill the criteria o f validity, must be reliable and the item test must be not too difficult and not too easy. These criteria of good test will be explained in chapter II.

Create a test is also crucial when it tests English language. Test writers do not only write a test just it is, but also try to conduct the test applicable as the way of language as communication tool. It is because as hughes said that language

[9] Arthur Hughes, *Testing for Language Teacher*, (Cambridge: University Press, 2003). Page 8

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

abilities are not easy to measure.[10] There are four major skills that should be carefully mastered in English language ability and used to perform as many genuinely communicative tasks as possible in situation where English is taught, they are listening, listening and speaking, reading and writing. [11] Thus, it is important to the test writer to concentrate on the items that relevant to the ability to use language for real-life communication.

Regarding to the case above, it is very important to have tests or some kind another, are valid, well designed and formulated. Hughes mentioned in his book that test is said to be valid if it is measure accurately what it should be measured. Nurkanca and Sumartana also pointed out that a qualified test should be reliable, valid and having degrees of difficulty-index and discriminating power.[12]

However, today there are a lot of concerned which are showed by means of mistrusting the tests by some educational experts and it frequently well founded. Moreover, nowadays there are many teachers do not careful in writing test. [13] They disregard the criteria of how to make a good test such as validity, reliability etc. Two researches done by student from UNESA in different school

---

[10] Ibid. Page 2

[11] J. B Heaton, *Writing English Language Test*, (New York: Longman Group, 1988). Page 8

[12] Wayan Nurkanca and Sumartana, *Evaluasi Pendidikan*, (Surabaya: Usaha Nasional, 1986). Page 127

[13] Layinatul Cholbi, The Final English Test for the fourth grade students of SDN Pucang III Sidoarjo, Unpublished S-1 Thesis (Surabaya: UNESA-FBS 2006), Page 2

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

have pointed out that the made by the teacher in each those schools are less reliability in essay tests and poor discriminate.[14]

In line of that case, the writer does the study, analyzing the content validity and item analysis on UAS Test for Tenth (X) Grade of SMAN 3 Sidoarjo. She needs to know whether the test is having good content validity and also having degrees of difficulty-index and discriminating power or not.

The writer chooses a test analysis for senior high school because she thinks that senior high school students should have a qualified teacher-made test. Moreover, senior high school is place where students prepare to continuo their study to university, thus the teacher-made test must be able to measure their progress and ability accurately.

The writer had read two previous theses which were focused on the same case. And both were done by students of English Department and Art of Surabaya State University (UNESA). The first was done by Laiyinatul Cholbi (2006). She analyzed The Final English Test for the fourth grade students of SDN Pucang III Sidoarjo. One of She conducted was the test had good agreement with the content of 1994 Local Content Curriculum for the fourth grade student of Elementary school (GBBP 1994). It means that the test had high content validity. The other Thesis was done by Ria Dhewi Pratiktasari (2006). She analyzed on semester II English summative test for the eighth year student of SMP Negeri 1 Slahung

[14] **Ria Dhewi Pratiktasari, An Analysis of Semester II English Summative Test for The Eighth Year Students of SMP Negeri I Slahung Ponorogo, Unpublished S-1 Thesis, (Surabaya: UNESA-FBS, 2006), Page 58.**

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Ponorogo. She found out that the test could not discriminate the upper level of student from the lower level. In the other word the index of discrimination of the test was poor.

Hence, the present study tried to analyze the test from the point of view of content validity, index of difficulty and also index of discrimination. The writer focused on The Final English Test for Tenth Grade of SMAN 3 Sidoarjo.

B. RESEACRH PROBLEMS

There are three problems that want to be revealed through this study, they are:

1. What is the UAS Test for tenth grade students of SMAN 3 Sidoarjo like in term of content validity?
2. What is the UAS Test for tenth grade students of SMAN 3 Sidoarjo like in term of Index of Difficulty?
3. What is the UAS Test for tenth grade students of SMAN 3 Sidoarjo like in term of Index of Discrimination?
4. What is the UAS Test for tenth grade students of SMAN 3 Sidoarjo like in term of the Effectiveness of Distractors?

C. PURPOSE OF THE STUDY

The study aimed to investigate the content validity based on standard competencies, index of difficulty, index of discrimination, and the Effectiveness of Distractors of UAS test for Tenth (X) grade of SMAN 3 Sidoarjo.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

D. SIGNIFICANT OF THE STUDY

This study is expected to give significant for:

1. The teacher: become an evaluation than later become a guideline to design better tests that matches the content validity.
2. The student: They get and are able to do the appropriate test questions that match the materials they got in the class, in order to measure their ability in mastering its materials.
3. The further researcher: As a references to do the others research.

E. SCOPE LIMITATION

The study only will concern on analyzing the agreement of the content validity and index of difficulty also index of discrimination of UAS test for Tenth (X) grade for SMAN 3 Sidoarjo. The writer will try to conduct whether test fulfills the criteria of good test from the point of view of content validity, index difficulty and also index of discrimination or it must be revised even replaced.

Furthermore, the study will take place in SMAN 3 Sidoarjo as the representative of senior high school around Sidoarjo since it is one of favorite school in Sidoarjo. It can be seen from many factors such as the teacher, the student and also the available facilities.

F. DEFINITION OF KEY TERM

To avoid misunderstanding in the way of understand the study, the writer need to give definition of terms that are often found in the study, such as:

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1. Content Validity:

Content validity refers to the test that representing all material course objectives that are being tested.[15]

Furthermore, what is being standard to measure the content validity in this thesis is based on *standar kompetensi* and *kompetensi dasar* which are stated in *standar isi* for based competence curriculum (Kurikulm berbasis competensi) of Tenth grade of SMA/MA in this case is SMAN 3 Sidoarjo.

2. Item Analysis

Item analysis means to indicate whether the test item is acceptable, less acceptable or poor. Item analysis purposes to get information about the weakness of a test and become references to revise the other next tests.[16]

3. Index of difficulty

Heaton stated that the index of difficulty of an item simply shows how easy or difficult the particular item proved in the test.[17]

---

[15] J. B Heaton, *Writing English Language Tests*, (New York: Longman Group, 1988), Page 160

[16] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta: Bina Aksara, 1984). Page 211

[17] J. B Heaton, *Writing English Language Test*, (New York: Longman Group, 1988). Page 178

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4. Index of discrimination

Heaton said that the index of discrimination indicates the extent to which the item discriminates between the testees, separating the more able testees from the less able[18]

5. Item Distractors

Item destractors means scrutinize the items great detail to know whether the distractors of items test performs as expected or not.[19] The good distractors will attract more poor students.

[18] *Ibid.* Page 179
[19] *Ibid.* Page 183

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

CHAPTER II

REVIEW OF RELATED LITERATURE

In this chapter, the writer will reviews several theories of literature related to the discussion of the study. It is presented to avoid misunderstanding to comprehend the study then. The subtitles that will be explained including definition of the test, types of the test, purpose of test, forms of test, characteristic of a good test and item analysis.

A. Definition of Test

Generally test is one type of measurement. Brown said that test is a set techniques, procedures, and items that constitute an instrument of some sort that requires performance or activity on the part of the test taker (and sometimes on the part of the tester as well)[1]

In line of that, test as quoted from Webster's Collegiate by Daryanto, is any series of questions or exercise or other means of measuring the skill, knowledge, intelligence, capacities of aptitudes or an individual or group.[2]

---

[1] H. Douglas Brown, *Teaching by Principles: An Interactive Approach To Language Pedagogy*, Second Edition, (San Francisco State University: Longman, Inc, 2001). Page 384

[2] Daryanto, Drs. H. *Evaluasi Pendidikan*, (Jakarta, Rineka Cipta, 1999). Page 35

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Furthermore, Muchtar Bukhori says:

> "Tes ialah suatu percobaan yang diadakan untuk mengetahui ada atau tidaknya hasil-hasil pelajaran tertentu pada seseorang murid atau kelompok murid"[3]

In the other word, Kubizyn and Borich stated in their book, that test is just as tools that can contribute importantly to the process of evaluating pupils, the curriculum, and the teaching method.[4]

Those above are several definitions about test created by some experts. Although these were written in different words or sentences, however it expressed the same meaning that test is one tool of process evaluating pupils, curriculum, and teaching method to measure the skill, the work of curriculum, and the successful of the teaching method. In addition to the previous explanation, as one type of measurement, a test is necessarily quantifies characteristics of individuals to explicit procedures.[5]

---

[3] Ibid. Page 35

[4] Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement* (Singapore, John Wiley & Sons, INC, 2003), Page 1

[5] Lyle F. Bachman. *Fundamental Consideration in Language Testing.* (New York, Oxford University Press, 1990), Page 20

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

B. Purpose of Test

As stated in the previous chapter that test has interrelated with teaching as well as education. Language tests also have many uses in educational programs, and sometimes two or more purposes cover the same test.[6]

David conducted six objectives of language testing:[7]

1. To determine readiness for instructional programs.
2. To classify or place individuals in appropriate language classes.
3. To diagnose the individual's specific strengths and weaknesses.
4. To measure aptitude for learning.
5. To measure the extent of student achievement of the instructional goals.
6. To evaluate the effectiveness of instruction.

In addition, Arikuto said that testing has several purposes for education, such as: a) testing is able to select the good student, b) testing is able to diagnosing the strength and weakness of student, c) testing is able to place students in proper class that fits their ability, and d) testing is able to measure the effectiveness of the program employed.[8]

---

[6] David P Harris, *Testing English as a second Language*, (New York: Mc Craw-ill, Inc, 1959), Page 2

[7] Ibid

[8] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta: Bumi Aksara, 1993), Page 9

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

In similar words, several purposes of test as cited from Hughes as follows:[9]

1. To measure language proficiency
2. To discover how successful students have been in achieving the objective of a course of study
3. To diagnose students' strengths and weaknesses, to identify what they know and what they do not know.
4. To assist placement of students b identifying the stage or part of a teaching programme most appropriate to their ability.

Ebel Also mentioned several benefits of test for both students and teacher/instractor,[10] such as, to measure students achievement and thus to contribute to the evaluation of educational progress and attainment. Test also benefits to motivate and direct students learning. In the generally fact, students tend to learn harder when they are examined or tested. They also stress to learn on the subjects that are tested. Not last, test can cause instructor/teacher to think carefully about the goals of instruction in a course.

---

[9] Arthur Hughes, *Testing for Language Teachers*, (Australia, Cambridge University Press, 1989), Page 8

[10] Robert L. Ebel, Essential of Educational Measurement. (USA, Prentice-Hall INC New Jersey, 1979). Page 22

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

C. Types of The Test

1. Based on the Number of test-taker

Based on the number of the test taker, test is divided into:[11]

a. Individual test: refers to a test where the tester tests only one testee, while

b. Group test: refers to a test where the taster faces more than one testee.

2. Based on the test maker

Test could be determined into teacher-made test and standard test. Although both are have the same purpose to measure the progress of teaching learning process, however they differ each other.

a. Teacher-made Test

Teacher made test is the test that is made the teacher of that classroom/or course itself. This test purposes to know measure how far the students achieve the instructional aim of particular lesson or course that are taught in the classroom.

Teacher made tests are designed based on the particular aim and description of the lesson that are taught in that class. Generally, this kind

[11] Prof. Drs. Anas Sudijono, Pengantar Evaluasi Pendidikan. (Jakarta: PT. Raja Grafindo Persada, 1996), Page 74

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

of test is not tried out before and even revised after. Thus, the validity of teacher made test often considered poor.[12]

b. Standardize Test

Standard test is a test constructed by test construction specialist, usually with the assistance of curriculum experts, teachers, and school administrators.[13]

Unlike teacher made test, standard tests are tried out before used. The tried out is not aimed to measure student's ability but to know whether the items test is adequate. The result of the tried out then analyzed through item analysis to get the coefficient of index difficulty and also index of discrimination. After that, the item that too difficult or too easy is should be revised as well as the item that has poor index of discrimination should be replace.

In addition, standard test is administered and scored according to specific and uniform. Thus it can be used in all and different school even it can used many times.[14] In the other word, a standard test administered

---

[12] Burhan Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra,* (Yogyakarta, BPFE-Yogyakarta, 2001), Page 60

[13] Tom Kubiszyn and Gary Borich, Educational Testing and Measurement (Singapore, John Wiley & Sons, INC, 2003), Page 343

[14] Ibid

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

and scored in Surabaya, would be administered and score in exactly the same manner in Malang, Sidoarjo, or anywhere in Indonesia.

3. Based on the purpose

a. Achievement Tests

Achievement tests are far more formal test and are intended to measure achievement on large scale. Brown said that an achievement test is related directly to classroom lesson, syllabus, or even a total curriculum.[15] He also added that achievement test is limited to specific material covered in curriculum within particular time frame, and offered after a course has covered the objectives question. In line of that, Heaton stated that this kind of test is design to measure ability based on what the students are predicted to have learnt, not necessarily on what they have actually learnt nor on what has been taught.[16]

Several achievement tests are standardized: they are pre-tested, each item is analyzed and revised where necessary, norms are established and comparisons made between performances of different students and different schools. A good achievement test should mirror the particular approach to learning and teaching that has previously been adopted

---

[15] H. Douglas Brown, *Teaching by Principles an Interactive approach to language pedagogy*, (San Francisco State University: Longman, Inc, 2001). Page 391

[16] J. B Heaton, *Writing English Language Test*, (New York: Longman Group, 1988). Page 172

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

In addition, Achievement test are divide into three types of test, they are:

1) Entry Behavior Test

Entry behavior test is held before student start to learn in the particular department of education (school). This kind of test is aimed to know whether the test taker provides abilities or skills which are required by particular school as accepting condition. It means those students who fulfill that criteria of accepting condition are able to accept in that school and vice versa.

2) Placement Test

Placement tests are used to give information that will help to put students at the stage of the teaching programm most suitable to their abilities.[17]

For instance, an English course has three level of classroom, elementary, intermediate and advance. In order to put the students in the proper class based on their ability, diagnostic test is held. The criteria of score in each class have been determined. Students who get particular determined score for elementary level are proper to place in

[17] Arthur Hughes, Testing for Language Teacher, ((Australia, Cambridge University Press, 1989), Page 17

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

that class. While the testees who achieve particular score for intermediate level is proper to put in the same class, and so advance.

This test is aimed to gather students which have at least similar level of ability in one classroom. It can help much to ease designing syllabus or planning classroom activities since the class has bit homogeny students. Thus, place students in wrong class or do not suit their ability must be pursue the effectiveness of teaching learning activities.

3) Diagnostic Test

Diagnostic test is used to identify learners' strengths and weakness.[18] Generally it is used primarily to determine what the learning still need to take place. The areas of difficulty are diagnosed in such of tests so that appropriate remedial action can be taken later.

To more focused and noted for teachers that the diagnostic test often held for group of students rather than for individuals. Thus, if only one or two students make a particular error, the teacher should not pay too much attention. Nevertheless, when a group of several students make a certain error, then the teacher will note the error and plan efficient remedial teaching.

[18] Ibid, Page 33

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

b. Formative Test

Formative test is held during the activities of teaching learning is going on. Usually it is done in the end of the accomplishment of one course. Thus, formative test might be held many times in one semester.

Formative test is aimed when teachers need to check on the progress of their students, to see how far they have achieved what they should have learned.[19] The information conducted from this kind of test enable teachers to measure the effectiveness of their classroom activities, and also to modify their future teaching plans.

c. Summative Test

Summative test is used at, say, the end of the term, semester or year. It is aimed to measure what has been achieved both group and individuals.[20]The materials that are tested including all course objectives which have been learned during one semester or year.

Unlike Formative test, summative test has more general purpose. The general purpose here have already stated clearly in the *standar isi* of every lesson.

---

[19] Arthur Hughes, *Testing for Language Teachers*, (Australia, Cambridge University Press, 1989), Page 5

[20]Arthur Hughes, Testing for Language Teachers, (Australia, Cambridge University Press, 1989), Page 5

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

The information from summative test is used to determine the score or achievement of each student. This information then planned to know the ability of one student among other, is he/she belong to upper group, middle, or lower group. This information also used to judge whether he/she passes the course and able to move to the next class or not.

d. Proficiency Tests

Whereas an achievement test looks back on what should have been learnt, the proficiency test is in the other hand, it looks forward and defining a student's language proficiency with reference to a particular task which he or she will be required to perform.[21] When the test is design to conduct global competence, then the test is in term of Proficiency test.[22]

Proficiency test is not intended to be limited to any one course, curriculum, or single skill in the language.[23] Proficiency test is usually created traditionally in form of multiple choice items on grammar, vocabulary, reading comprehension, and sometimes a sample of writing.

---

[21] Ibid. Page 5

[22] H. Douglas Brown, *Teaching by Principles an Interactive approach to language pedagogy*, (San Francisco State University: Longman, Inc, 2001). Page 390

[23] Ibid. Page 390

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

e. Aptitude tests.

An aptitude test serves to indicate an individual's facility for acquiring specific skills and learning.[24]

Language learning aptitude is a complex matter, consisting of such factors as intelligence, age, motivation, memory, phonological sensitive and sensitivity to grammatical patterning.[25]

Aptitude tests generally seek to predict the student's probable strength and weakness in learning a foreign language by measuring performance in an artificial language.

D. Forms of Test

There are two kinds of form of test: objective and subjective test. The distinction between both tests is concern on method of scoring, and nothing else.[26] The following explanation will clarify enough about them.

1. Objective Test

Sudijono claimed that objective test is one type of test that is created using items tests, then what the entire test taker has to do is just answering the

---

[24] David P Harris, *Testing English as a second Language*, (New York: Mc Craw-ill, Inc, 1959), Page 3

[25] J. B Heaton, *Writing English Language Test*, (New York: Longman Group, 1988). Page 173

[26] Ibid, Page 22

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

question by choosing one among several probably answers available in each items or writing sentences or particular symbols in place provided in each item test.[27]

In line of that, objectives test as cited from Lado is:

"Objectives test are those that are scored rather than mechanically without need to evaluate complex performance on scale"[28]

a. Types of an objectives test

Sudijono also added that there are five types of objectives test including: true or false test, matching test, completion test, fill in test and also multiple choices. However in this thesis only will clarify the last one.

Multiple choices as stated by Sudijono are a test which is created likely incomplete sentences and the testee should complete the sentence in order to answer the question.[29]Before going to design multiple choice test, the test maker or in this case is teacher should know primarily several terms used in multiple choices. First is *stem* which refers to initial part of each multiple choice items. Second is option/responses/alternatives, refers to the options which are available for student to select their answer. One

---

[27] Prof. Drs. Anas Sudijono, *Pengantar Evaluasi Pendidikan*, (Jakarta: PT. Rayagrafindo Pustaka, 1990), Page 106

[28] Robert Lado, *Language Testing*, (London: Longman Group, 1961), Page 28

[29] Prof. Drs. Anas Sudijono, *Pengantar evaluasi Pendidikan*, (Jakarta: PT. RajaGrafindo Pustaka, 1996), Page 106

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

option among them is called *the correct answer* and the other is *distractors*[30]

The illustration from the explanation above as follows:[31]

Stay here until Mr. Short.....you to come. = *stem*

A. Told
B. will tell } = *options/response/alternatives* } = *distractors*
C. is telling
D. tells = *correct answer*

b. The Benefits and Weakness of Objective Tests

1) The Benefits of Objective Test

When objective test of language are properly made, they have important values. Arikunto mentioned several goodness of objective test:[32]

- Represent more all objective materials that are being tested
- They can test in short time
- They can be scored with speed and ease

---

[30] J. B. Heaton, *Writing English Language Tests*, (New York: Longman Group, 1988). Page 28

[31] Ibid

[32] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 164

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

- They use careful objective score in evaluating the test.
- They able to be scored not only by teacher or test maker.

2) The weakness of Objective Test

The usual objections to objective test are mentioned by Lado as follows:[33]

- They are too simple
- They do not require real thinking but simply memory
- They do not test the ability of the students to organize his thoughts.

Beside the three previous objective test's weakness, Arikunto also added the rest objections[34]:

- Objective test enable student being speculative in responding the question in a test
- Open widely possibilities for students to cheating each other in doing a test
- It is more difficult to construct the objectives test than subjective test because it contains a lot of item tests.

[33] Robert Lado, *Language Testing*, (London: Longman Group, 1961), Page 35
[34] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 164

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. Subjective Test

As quoted from Lado, subjective test is:

"Tests that require an opinion and a judgment on the part of the examiner"[35].

In the other word, Nurgiyantoro have said that subjective test is a test that require student to answer in essay using their word.[36]

a. Scoring an essay test

Scoring an essay test generally based on the weight of each item test, the level of difficulty, and the amount of the element contained by the answer which is considered as the rightes answer.

For example, there are 5 items test in essay test. The tester had determined that all items have the same level of difficulty, and the elements in each item had made in the same amount. Based on that, tester decided that testee who could answer with the rightest answer or which the answer provides the entire element that required by the tester within the item test, will get 10 marks. When the testee answer almost perfectly or the answer provide mostly the element that required by the test taker, will get 9 mark, and so on.[37]

---

[35] Robert Lado, *Language Testing*, (London: Longman Group, 1961), Page 28

[36] Burhan Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta, BPFE Yogyakarta, 2001), Page 71

[37] Prof. Drs. Anas Sudijono, *Pengantar Evaluasi Pendidikan*, (Jakarta, PT. Raja Grafindo persada, 1996), Page 301

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1). **The Benefit and the weakness of subjective test**

The characteristics of subjective can be seen from its benefits and weakness as follow

a). The Benefit

- Subjective test can create easily and fast
- Avoid students being speculative in answering the items test
- The test taker is able to know how far students understand the material
- Motivate student to organize their thoughts

b). The weakness

- Less able to represent all materials
- It is difficult to score the subjective test. It because the answer of each item might be varieties and wide. Thus, it needs a lot of time, and thoughts to score it.
- Enable test taker to score subjectively
- Validity and reliability of subjective test is poor.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

E. Characteristic of A Good Test

All good tests possess three qualities: validity, reliability, and practicality.[38] In the other word we say, any test that we use must be appropriate in terms of our objectives (validity), dependable in the evidence it provides (reliability), and applicable to our particular situation (practicality). Without any one of them, a test would be a poor investment in time and money.

Whether the teacher is constructing his own test or selecting a standard instrument for use in his class or school, he should certainly understand what these concepts mean and how to apply them.

1. Validity

Validity of a test is the extent to which it measures what it is supposed to measure.[39] In different word but still have same meaning; Gronlund said that validity refers to the appropriateness of interpretations made from test scores and other evaluation result, with regard to a particular use.[40]

Furthermore, a test has validity evidence if we can demonstrate that it measures what it says it measures. For example, if the test is supposed to be a test of tenth-grade English language ability, it should measure tenth-grade

---

[38] David P Harris, *Testing English as a second Language*, (New York: Mc Craw-ill, Inc, 1959), Page 159

[39] J.B Heaton, *Writing English Language Test*, (New York: Longman Group, 1988). Page 172

[40] Norman E. Grounlund, *Measurement and Evaluation in Teaching*, (New York, collier Macmillan Publisher, 1985), page 55

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

skill, not eleventh-grade skill and not mathematic ability. If it is supposed to be a measure reading ability, then it should measure reading ability, not the ability of writing, and so on. To conclude, we can simply say that validity describes how well a test or evaluative technique does the job that it is employed to do.[41]

In addition, there are several ways of deciding whether a test has sufficient validity evidence:

a. Content Validity

The simplest way to judge the validity of a test is content validity evidence. Content validity is the process of determining the extent to which a set of test tasks provides a relevant and representative sample of the domain of tasks under consideration.[42]In line of that, content validity as quoted from Thoha, is

> "Validitas isi mempersoalkan apakah isi butir tes yang diujikan itu mencerminkan isi kurikulum yang seharusnya diukur atau tidak"[43]

In the easier word we can say that content validity describe how well the samples of test tasks represent the domain of tasks to

---

[41] Drs. M. Ngalim Purwanto, *Prinsip-prinsip dan Teknik evaluasi Pengajaran*, (Bandung, Remadja Karya, 1985, Page137)

[42] Norman E. Grounlund, *Measurement and Evaluation in Teaching*, (New York, collier Macmillan Publisher, 1985), page 59

[43] . M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, (Jakarta, PT RajaGrafindo Persada, 1991), Page 111

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

measure.[44]Furthermore, the content validity evidence for a test is established by examination. Test questions are inspected to see whether they correspond to what the user decides should be covered by the test.[45] For instance, when the test is supposed to test some particular objectives of arithmetic material that are taught in third-grade, but in the real it tests out of those objectives or in the other word, it tests other material objectives of arithmetic which are not taught in that class, then the test is called has poor content validity.

However, a test can sometimes look valid but measure something entirely different than what is intended. Content validity is, therefore, more a minimum requirement for a useful test than it is a guarantee of a good test.

In summary, content validity evidence answers the question"Does the test measures the instructional objectives?" In the other word, a test with good content validity evidence matches or fits the instructional objectives.[46]

---

[44] Ibid, Page 58.

[45] Tom Kubiszyn and Gary Borich, Educational Testing and Measurement (Singapore, John Wiley & Sons, INC, 2003), Page 300

[46] Ibid

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

b. Criterion-Related Validity

A second form of validity evidence is criterion-related validity evidence. Criterion-related validity evidence is refers when the scores from a test are correlated with an external criterion.[47]

In addition, there are two types of criterion-related validity evidence: concurrent and predictive.

1) Concurrent Validity

Concurrent validity is deals with the relationship between two sets of test scores. Kubiszyn said that concurrent validity evidence is refers to the measures that can be administered at the same time as the measure to be valid.[48]

For example: A teacher who wants to measure the validity of the test score of Islamic education in one of senior high school. Then he compares the score of that test with the score from oral assessment and observation assessment. When he scores from both assessment has positive correlation, then the test is called having good concurrent validity evidence because it expressed correlation coefficient which means two different assessments measure the same thing in the same

[47] Tom Kubiszyn and Gary Borich, Educational Testing and Measurement (Singapore, John Wiley & Sons, INC, 2003), Page 300
[48] Ibid

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

time and has the same relative score. Thus, in the other word, concurrent validity evidence as quoted from Toha, is:

"Validitas pengukuran artinya test tersebur memiliki kesesuaian dengan hasil pengukuran yang lain yang dilaksanakan pada saat itu juga dengan menggunakan alat ukur yang berbeda"

2) Predictive validity

Predictive validity refers to the extent of the test can predict the examinee's aptitude in the future.[49] In the other word but still has the same meaning, Hughes proclaimed that the test is claimed having predictive validity evidence when the test can predict the candidates' future performance.[50]

c. Face Validity

Face validity is related to the appearance of the test. Hughes said that a test is called has face validity when it pictures as if it measures what it is employed to.[51]For example, a test that created to measure pronunciation ability but which did not require the testee to speak might be considered to lack face validity. The test is said to have face validity also require good performance of the test itself including the quality of the paper used, the

[49] Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement* (Singapore, John Wiley & Sons, INC, 2003), Page 301

[50] Arthur Hughes, *Testing for Language Teachers*, (Australia, Cambridge University Press, 1989), Page 29

[51] Ibid, Page 33

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

pictures that are representative of the objective materials, and also the quality of the color.

Furthermore, a test that has poor face validity is not providing evidence to judge that it is lack of validity, because face validity is not a scientific notion.[52]Nevertheless, face validity is very important to be considered of writing tests. A test which does not have face validity may not be accepted by the testee, teachers, education authorities or employers.

d. Construct validity

A test has construct validity evidence if its relationship to other information corresponds well with some theory.[53]A theory is simply a logical explanation or rationale that can account for the interrelationship among a set of variables.

Construct validity evidence is more specific and immediately practical uses than the others, we may wish to interpret test scores in terms of their psychological meaning.[54]

---

[52] Ibid. Page 33

[53] Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement* (Singapore, John Wiley & Sons, INC, 2003), Page 302

[54] Norman E. Grounlund, *Measurement and Evaluation in Teaching*, (New York, collier Macmillan Publisher, 1985), page 72

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

In the other word, construct validity as quoted from Thoha, is:

> "Tes yang butir-butir soalnya mengukur aspek berfikir sesuai dengan konsep atau pendekatan yang digunakan untuk mengurai aspek berfikir tersebut"[55]

2. Reliability

Another powerful evidence to judge whether the test is called good test is through reliability. Reliability refers to the stability of the test in presenting the score. Harry added that a test cannot measure anything unless it measure consistently.[56]

3. Practicality

Brown stated that a good test must be practical. It is within the means of financial limitations, time constraints, case of administration, and scoring and interpretation. Thus, he said that a test which is prohibitively expensive is impractical. A test proficiency that takes a student ten hours to complete is impractical. And also a test that takes few minutes for a student to take and several hours for an examiner to evaluate is impractical for most common

---

[55] Drs. M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, (Jakarta, PT RajaGrafindo Persada, 1991), Page 110

[56] David P Harris, *Testing English as a second Language*, (New York: Mc Craw-ill, Inc, 1959), Page 14

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

classroom. The value and quality of a test are dependent upon such practical considerations[57]

F. Item Analysis

The most effective way to evaluate the teaching learning process is evaluating the test within the teaching learning process it selves. In the other word, we treat or evaluate the score of the test in order to know and place which component from the teaching learning process is weak or need to be revised.

Purwanto said that there are two ways to evaluate test, such as: item analysis and evaluating the validity and the reliability of the test.[58] However, the following explanations only will focus on first way since the second have already stated in the previous sub items

When the test considered as a good test? Arikunto answered that tests are acceptable if it matches three kinds of item analysis, such as: it difficulty, item discrimination, and also distractor.[59]

---

[57] H. Douglas Brown, *Teaching by Principles: An Interactive Approach To Language Pedagogy*, Second Edition, (San Francisco State University: Longman, Inc, 2001). Page 386.

[58] Drs. M. Ngalim Purwanto, *Prinsip-prinsip dan Teknik Evaluasi Pengajaran*. (Bandung, PT. Remaja Rosdakarya, 1984), Page118

[59] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 209

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1. Index of Difficulty

Good test is not too easy and too difficult either. Test which is too easy is not motivating students to try hardly answering the test. In cross way, test which is too difficult causes student to be suffer finishing the test beyond what they can reach out.

The index of difficulty of an item simply shows how easy or difficult the particular item proved in the test.[60]The index of difficulty is generally expressed the percentage of students who answer correctly. Arikunto use the following formula to calculated index of difficulty[61]:

$$P = \frac{B}{JS}$$

Where:

P = Index of difficulty

B = the number of students who answer correctly

JS = the number if students who taking the test

.Nevertheless, Heaton used another formula in different terms but still have the same meaning as follows:

---

[60] J. B Heaton, *Writing English Language Test*, (New York: Longman Group, 1988). Page 178

[61] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 210

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$FV = \frac{R}{N}$$

R represents the number of correct answer and N is the number of students taking a test.

Amount the number of index difficulty between 00 to 1,0. These numbers express the difficulty level of an item test. Some experts are difference in giving the amount of number to express the level of difficulty of an item. Oller in Burhan stated that an item test which has index of difficulty between 0,15 up to 0, 85 is adequate, out of that numbers is too easy and too difficult. Thus it is needed to be revised or changed.[62] However, Arikunto given a common ukuran about the number of index of difficulty as follow[63]:

- The item test with index of difficulty 00 up to 0,30 means it is difficult
- The item test with index of difficulty 0, 30 up to 0, 70 means it is good
- The item test with index of difficulty 0,70 up to 1,00 means it is easy

[62] Burhanudin Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta, BPFE Yogyakarta, 2001)Page 138

[63] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 212

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. Index of Discrimination

Index of discrimination of an item is the ability of item test to differentiate the up level student from down level.[64]

The number which shows amount of index of discrimination is called index of discrimination. Like index of difficulty, the number of index of discrimination is between 00 up to 1, 00. However, unlike index of difficulty, index of discrimination has (-) negative sign. It is used when test discriminate in entirely in wrong way showing the quality or ability of testee.[65] It shows when none of upper level students got a correct answer and the lower level student can answer correctly.

In addition, Heaton stated several ways in analyzing index of discrimination of test item:

1. Arrange the script in rank order of total score and divide into two groups of equal size (the top half and the bottom half). If there is an odd number of script, dispense with one script chosen at random
2. Count the number of those candidate in the upper group answering the first item correctly, then count the number of those candidate in lower level group candidates answering the item correctly, and so on.
3. Subtract the number of correct answers in the upper level group from the number of correct answer in lower level group. Find the difference in the

---

[64] Ibid, Page 213

[65] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 213

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

KEMENTERIAN AGAMA PERPUSTAKAAN SURABAYA

proportion passing in the upper group and the proportion in passing in the lower group.

4. Divide this difference by the total number of candidates in one group:

$$D = \frac{\text{Correct U} - \text{Correct L}}{n}$$

(D = Discrimination index, n = Number of candidates in one group, U = Upper level half and L = Lower level half. The index D is the difference between the proportion passing the item in U and L)

5. Proceed in this manner for each item.

Furthermore, Arikunto claimed that items test is called good when it has index of discrimination in range of 0,40 up to 0,70. He also added the classification of the number range of index of discrimination of an item test as follow[66]:

D : 00 up to o,20 means it is poor

D : 0,20 up to 0,40 means it is satisfactory

D : 0,40 up to 0,70 means it is good

D : 0,70 up to 1,00 means it is excellent

3. Analysis Distractor

Analysis distractor means analyzing on the distributing of test taker in determining the option of answering the question in the multiple choice test. [67] Analyzing distractor conducted through calculating the amount of test taker

[66] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 221

[67] Ibid, Page 225

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

who choose the option a, b, c, d, or e, or test taker who do choose none of the option. In this case, called omit (O).

Furthermore, information conducted from analyzing distractor planned to know whether the distractor play good part in the option or not. Distractors must attract more students in lower group. Therefore, if the distructors chosen by more able students, it means they are poor. Besides, distractor which is not choosen by all students shows that they cannot perform well, thus all alternatives must be selected by the test taker. Moreover, when there is the same amount of the voter from both better and poorer students who chose those distractors, means they are still desirable adequate to be used for future test. But when there is divergence of those matters above, the distractors are suggested to be revised and cannot be applied for others test.[68]

In addition, distractors called to be good distractor if it is chosen by at least 5% of students. It also called effective if the omit is chosen by not more than 10%.[69]

---

[68] Burhan Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta, BPFE Yogyakarta, 2001). Page 144.

[69] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 226

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

CHAPTER III

RESEARCH METHODOLOGY

In this chapter will employ the research methodology. Means here the writer will formulate the research design that used by the writer in the way of analyzing the study, they are: Research design, Data and Source of Data, Data Collection, Instrument of Data Collection and Data Analysis and Schedule of Research.

A. Research Design

Mardalis categorized four types of research method which are often used; they are Historical research, explorative research, descriptive research and explanatory research.[1] Descriptive research is used to obtain information concerning the current status of the phenomena to describe "what exists" with respect to variables or conditions in a situation. The methods involved range from the survey which describes the status quo, the correlation study which investigates the relationship between variables, to developmental studies which seek to determine changes over time.[2]

Based on the statement above, the study entitled "Content Validity and Item Analysis on UAS Test for Tenth (X) Grade of SMAN 3 Sidoarjo", used descriptive research as the way or technique to do the research since it will describe the validity of the content and will also describe the index of difficulty and index of discrimination on UAS test for Tenth (X) Grade of

[1] Drs. Mardalis, *Metode Penelitian*, (Jakarta: Bumi Aksara,1995). Page 25
[2] *Ibid.* Page 25

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

SMAN 3 Sidoarjo. Besides, this study also used quantitative approach since it is used numerical calculation to compute the data.

B. Research Subject

Furthermore, the subject of this study is semester II English Final Test for tenth grade (X) of SMAN 3 Sidoarjo and takes place in X1, X2, and X3 class.

In X1 class there are 36 students, in X2 class there are 36 students andin X3 class there are 35 students. The writer used random sample technique to determine the sample. Furthermore, the sample from this population is used to help in analyzing the test item (Item analysis).

C. Data and Source of Data

According to Arikunto, the source of data is the place or things in which the researcher can observe, ask or read about related matter of the object being studied. It can be divided into person, place and documentation.[3]

Furthermore, in this study the data will be obtained from the items of UAS test, the key answer, the student's answer sheets and also the result of the student's score and the standard and basic competencies of 2006 English curriculum for tenth grade of senior high school.

It will be used to conduct the data about the following points:

1. The test items and key answers of The Second Semester Final English Test (2009/2010) for the Tenth grade student of SMAN 3 Sidoarjo.

[3] Suharsimi Arikunto, *Prosedur Penelitian*, (Jakarta: PT. Rineka Cipta, 1996) P. 123

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. The student's answer sheets of The Second Semester Final English Test (2009/2010) for the Tenth grade student of SMAN 3 Sidoarjo
3. The student's score of the Second Semester Final English Test (2009-2010)
4. The standard and basic competencies of 2006 English curriculum for tenth grade of senior high school.

D. Data Collection Techniques

The data in this study will be collected by study documentation. Documentation is a method to get anything on the form of notes, transcripts, magazines, books, etc,[4] and then data by documentation will be collected through these following steps:

1. Finding the test items, key answer, the student's answer sheets, the teaching materials and student's score of the final test.
2. Finding the Standard and Basic Competencies of the 2006 English curriculum for the tenth grade of senior high school

E. Instrument of Data Collection

According to Mardalis instrument by means of researching is the implement measured. That is with instrument this research could be gathered by the data as the implement to state the mulberry or the percentage as well as more the shortage in the form of quantitative or qualitative[5]

---

[4] *Ibid.* Page 234
[5] Drs. Mardalis, *Metode Penelitian*, (Jakarta: Bumi Aksara,1995). Page 60

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

This study used study documentation to measure the validity of collected data and also interview the teachers to support the data.

F. Data Analysis

The result of collected data then will be analyses by using descriptive, means that data will be described as the way it is.

1. Analyzing The Content Validity

In analyzing the content validity, the writer will collect it through the following steps:

a. Making a list of the standard competencies, basic competencies, indicators, and learning experience for the tenth grade students of senior high school and the indicators of basic competencies given by SMAN 3 for tenth grade student.

b. Placing each of the test items in the appropriate place with the standard competencies and basic competencies to identify whether or not the standard competencies and basic competencies covered by the final test.

c. Counting the percentage of the test items of every language aspects.

d. Concluding the result of analysis.

In order to make these procedures clearer, the writer presents the illustration of the procedures as follows:

Table 1. The example of analyzing Content Validity

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

STANDARD COMPETENCE	BASIC COMPETENCE	INDICATORS	Learning Experience	ITEM TEST	Σ	%
NDENGARKAN mahami makna am percakapan saksional dan rpersonal dalam tek kehidupan ari-hari	Merespon makna dalam percakapan transaksional (to get thing done) dan interpersonal (bersosialisasi) resmi dan tak resmi secara akurat, lnacar dan berterima yang menggunakan ragam bahasa lisan sederhana dalam berbagai kontek kehidupan sehari-hari dan melibatkan tindak tutur: berterima kasih, memuji, dan mengucapkan selamat	❖ Mendidentifikasi kata yang didiengar, makna kata, hubungan antar pembicara,. ❖ Mengidentifikasi makna tindak tutur, berterimakasih, memuji, mengucapkan salam dan konek situasi. ❖ Mereposn tindak tutur, berterimakasih, memuji, ucapan, selamat	• Mendengarkan berbagai tindak tutur yang didengar melalui tape atau teman • Mendiskusikan berbagai tindak tutur yang didengar melalui tape atau teman	2,5,10,7, 55, 4, 6, 8, 11, 20, 21, 24, 28, 30, 34, 35	5 11	9% 20%

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. Analyzing the Index of Difficulty

The Index of difficulty of an item simply shows how easy or difficult the particular item proved in the test.[6]To analyze the index of difficulty of test items, the writer takes the following steps:

a. Arranging the students' score from the highest score to the lowest one.

b. Finding the top and the bottom of the students' score, as upper and lower groups. Dividing the scripts in rank order of total score into two groups of equal size, the top half as the upper level and the bottom half as the lower group.

c. Computing the item difficulty by using the formula of by Heaton below:[7]

$$FV = \frac{R}{N}$$

Where:

FV = Index of difficulty

R = the number of students who answer correctly

N = the number of students who taking the test

[6] J. B Heaton, Writing English Language Test, (New York: Longman Group, 1988). Page 178

[7] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 210

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

3. Analyzing the Index of Discrimination

The Index of Discrimination indicates the extent to which the item discriminates between the testees, separating the more able testees from the less able.[8] To analyze the Index of Discrimination here use the same steps using in analyze Index of difficulty. Then, to calculate the index of discrimination, the writer used the formula below:[9]

$$D = \frac{\text{Correct U} - \text{Correct L}}{n}$$

Where

D : The Index of Discrimination

Correct U : The number of students in upper group who answered the items correctly

Correct L : The number of students in lower group who answer the items correctly

n : The number of students taking the test in one group.

4. Analyzing the Effectiveness of Distractor

Besides calculating index of difficulty and discrimination, it also important to analyze the items in very detail, moreover on those which cannot perform as expected. Analyzing the distractor aimed not only to

[8] J. B Heaton, Writing English Language Test, (New York: Longman Group, 1988). Page 179
[9] *Ibid.* Page 180

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

know which items that cannot work properly but also to check why particular test taker failed to answer certain items correctly.

Distractors shave functioned well if these chosen mostly by students from lower level. According to Arikunto, the distractor which is chosen at least by 5% students from is called good distractor.

In addition, to conduct the effectiveness of distractor the writer should determine the amount of students from upper and lower level who chosen each options in each item. The writer also determines the amount of students who do not chose the options at all (omit). However, to ease the analyzing, the writer used a table as follow:

Table 2. The example of analyzing the Effectiveness of Distractors[10]

Item Number	Options	Upper Group	Lower Group	Comment
1	A	1	8	Good
	B*	22	11	Good
	C	1	2	Good

[10] Layinatul Cholbi, *The Final English Test for the fourth grade students of SDN Pucang III Sidoarjo,* Unpublished S-1 Thesis (Surabaya: UNESA-FBS 2006), Page 31

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	D O	1 0	4 0	Good NF

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

CHAPTER IV

DATA ANALYSIS

In this chapter the writer will discuss the data analysis. In this section, the writer will present the process of analyzing the data and also discuss the result of data analysis to get information about content validity, index of difficulty, index of discrimination, and also the effectiveness of destructor of semester II English Final Test for tenth grade of SMAN 3 Sidoarjo.

Furthermore, all data that helped writer to analyzed were conducted from documents belong to first grade of SMAN 3, such as, the item test, the answer keys, the students answer sheet, and the students' score.

A. Analyzing Content Validity

In analyzing content validity, the *Standar Isi* table including standard competencies and basic competencies of 2006 curriculum is used. Then, to judge whether the Semester II English Final Test for Tenth grade student of SMAN 3 Sidoarjo fulfill the agreement of content validity or not, the *standar Isi* table is matched with the materials in the test. Furthermore, place each item number in appropriate intersection of *standar isi* table of 2006 curriculum to identify the representative sample.

Table analysis in the table 3 of appendix I is used to ease the writer in analyzing content validity of Semester II English final Test for Tenth Grade of

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

SMAN 3 Sidoarjo. There are 6 columns in that table. First column contains standard competence, second column contains basic competencies, the third column contains several indicators that represent the basic competence of the lesson, and the fourth column contains the learning experience or materials that are taught. While the fourth, fifth and the last column are contain the items test that appropriate the basic competence, the sum of the items test that appropriate the basic competence and the percentage of total numbers of particular items that represent the related basic competence.

In addition, according to Nurgiyantoro, the test has the content validity if it covers all the contents as stated in the curriculum. Based on the result of analyzing content validity in table 3 appendix I, the percentage of every aspect of the learning content is conclude as follows:

1. There are 3 or 5% items for speaking which focus on direct and indirect speech explanation

2. There are 3 or 5% items for speaking which focus on passive voice explanation

3. There are 35 or 63% items for reading skill out of 55, which 2 items or 3 % focus on " Membaca dan memahami pengumuman/surat" lesson, one or 1% focus on " Memahami makna teks yang dibaca", 18 items or 32% focus on "Memahami isi teks yang dibaca", 3 items or 5% focus on "Memperhatikan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dan menemukan cirri-ciri kebahasaan teks" lesson, 4 items or 7% focus on "melengkapi teks dengan kata kerja yang tepat" lerning experience, and 7 items or 12% focus on "Menentukan orientasi cerita dengan metode yang berbeda" Learning experience

4. There are also 14 items or 25% out of 55 items test that did not cover the available materials.

Based on the result above, we can conclude that the Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo is good since 72% items test represented all materials. Here the agreement of the test is more than 50%. According to Bloom, if the agreement of the test is 50% or more, it can be conclude that the test has high content validity. [1]

Nevertheless, there are still 14 items test or 25% out of all items test did not cover the materials, they items number 8, 12, 17, 18, 20, 21, 22, 25, 45, 48, 49, 52, 53, and 54. Although their content is suitable with the indicators of standard and basic competencies but they were not taught in the class. Those items cannot be tested because students absolutely cannot answer that items test.

[1] Benjamin S. Bloom, at all, *Evaluation to Improve Learning,(USA, 1981)*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

B. Analyzing Index of difficulty

To conduct index difficulty value, divide the number of students got the correct answer by the number of students taking test.

In analysing the index of difficulty, first of all, the writer arranged the students' score from the highest score to the lowest one. Then the writer found the top score and the bottom score and divided it into two groups, upper level group and lower level group of equal size.

In addition, the writer treated differently in dividing class into two groups in order to determine upper and lower level in each class in the same amount moreover in X1 and X3 classes. This is because there were odd numbers of script in those classes.

In X1 class there are 3 students belong to upper group who got the same score, 24. It made difficult in dividing the class into two groups in equal size. It should have been 18 students for upper group and 18 students for lower group since the class has 36 students in all, but it became 21 students for upper group and 15 students for lower group. Hence, the writer deleted two number of script in random in order to make it balance as Heaton stated in his book:

> "If there is an odd number of a script, dispense with one script chosen in random"[2]

[2] J. B Heaton, *Writing English Language Test*, (New York: Longman Group, 1988). Page 178

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Although Heaton suggested dispensing with one script if there is an odd number, but in this case the writer deleted two numbers at all. It is because if only one script was deleted, it still does not make sense since the odd number existed in the same order.

Thus, the number of students taking test in X1 that used to compute index of difficulty and index of discrimination later are 34 students. While in X3 class, the writer dispensed with one script since this class has odd number of script. Like stated above, X3 class has 35 number of student taking test. Hence, in this case, the number of test takers that used to calculating index off difficulty as well as index of discrimination is 34 students.

After determining the upper and lower group of students, the writer computed the index of difficulty using Heaton formula as stated in Chapter II and III. Furthermore, the writer used a table to make calculation easy and efficient. The table can be seen in appendix 5-7.

There are six columns in the table. First column contains the number of item. Second column contains the score of students in upper group who answer correctly of each item. Third column contains the score of students in lower group who answer correctly of each item. The fourth and fifth column contains the value of index of difficulty Index and the value of index of discrimination. And the last

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

column contains the comment. The column of comment divided into comment for Index difficulty value and for index of discrimination value.

However, the writer calculated index of difficulty value of items test for tenth grade of SMAN 3 Sidoarjo. The writer computed the item test for three classes since this thesis used three samples such as X1, X2, and X3. The writer started calculating from X1, X2, and the last X3.

1. Analyzing Index of Difficulty on UAS Test for X1 Class

The result of analysis shown in table 7 appendix 5, reported that there are 27 out of 55 items or 13. 25% of item test have index of difficulty value between 0.32-0.65, they are items number 3, 5, 6, 9, 10, 11, 14, 15, 16, 17, 19, 21, 24, 25 26, 28, 30, 34, 37, 38, 39, 40, 43, 45, 46, 48, and 52. According to Arikunto, the items that show index of difficulty value between 0.30-0.70 are good[3]. It means that those are categorized adequate items and could be safety used in the future tests without being rewritten. Besides, the items number 12, 18, 22, 23, 29, 31, 32, 33, 35, 36, 47, 49, 50, 51, 53, 54, and 55 are difficult items test since they show difficulty value between 0.03-0.29, as Arikunto said that the items which have index of difficulty value between 0-00-0.30 are difficult item test. Thus, they are needed to be revised because they might desperate students to study more. While, the rest items are easy, such item

[3] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 212

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

number 1, 2, 4, 7, 8, 13, 20, 27, 41, 42, and 44 because they have index of difficulty value between 0.71-091. These items cannot to use for other test unless they are revised. It is because easy items test

2. Analyzing Index of Difficulty analysis on UAS Test for X2 Class

The table 8 of appendix 6 shown that there are 19 items test of X2 class have difficulty value between 0.72 up to 0.97such item number 1, 2, 3, 4, 7, 8, 9, 11, 13, 20, 21, 26, 27, 38, 39, 41, 43, 44, 45, and 48. These items test considered as easy item. Moreover, there is one item that categorized very easy since it has facility value more than 1.00, it is 1.14. Both easy items and very easy items are might be revised because they do not require students to organize their thought. All students whether good or bad might simply answer the items without spend hard think. Besides, the items test number 5, 6, 10, 14, 15, 16, 17, 18, 19, 24, 25, 28, 30, 31, 34, 36, 37, 40, 42, 47, 49, 53, 54, and 55 or 24 items are adequate because they have facility value between 0.31-0.69. It is relief since those items are desirable for applying in the following any tests because they are not need to be refresh. While 11 items the rest are difficult because they have difficulty value between 0.03-0.28. They are items number 12, 22, 23, 39, 32, 33, 35, 46, 50, 51, and 52. Those difficulty items that are badly needed to be rewritten because they cannot work as expected.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

3. Analyzing Index of Difficulty on UAS Test for X3 Class

From the result of calculating index of difficulty of test item for X3 shown in the table 9 appendix 7, there are 21 items that have index difficulty value between 0.74-0.91. These items are item number 1, 2, 4, 6, 7, 8, 9, 11, 17, 18, 20, 21, 26, 38, 39, 40, 41, 43, 45, 48. According to Arikunto the item that shows value between 0.70-1.00 is called easy[4]. It overcomes sad result since those easy items cannot function properly and cannot be used again before they are revised. Besides, there are 15 items are difficult since these have index of difficulty value between 0.00-0.29. There items number 12, 19, 22, 23, 29, 31, 32, 33, 34, 37, 50, 51, 52, and 54. Arikunto claimed that item which has index of difficulty value between 0.00-0.30 called difficult. Those items are needed to ministered and revised better because they are not able to play well. In addition, items number 3, 5, 10, 13, 14, 15, 16, 24, 25, 27, 28, 30, 35, 36, 42, 46, 47, 49, 53, 55 or it can be said 19 items show index of difficulty value between 0.35-0.68. These items are claimed as adequate item test since these items have index of difficulty between 0.30-070 as Arikunto stated in his book. It is relief since those items test are categorized good item. It means those items can be fully utilized and become reference to design other items for other tests.

---

[4] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 212

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Finally, after discussing the result of index difficulty analysis in each X1, X2, and X3 class, Generally it can be conclude that most of items test which used in those classes are good since they have facility value around 0.30-0.70 based on Arikunto's classification[5], although the result in X3 shown that most of items categorized easy because the difference value of easy items and good items only 1 number. Those good items may be used for the next test or able to be references for the test maker to design other test.

Nevertheless, there are still huge numbers of items test that shown as easy and difficult items. However, both that difficult and easy items must be revised because they cannot work as expected to measure student's progress and which students that able to absorb the lesson well. Difficult items test could make students confuse and spend most of their time only focus to answer those certain items. They will not care about the time available for doing the test. At the end, students will use their last minute to answer as fast as possible the rest items test desperately. In the other hand, the items test that easy do not give chance to students to organized their thought, because whoever can answer those items easily without spend hard work.

[5] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 212

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

C. Analyzing Index of Discrimination

To conduct the discrimination value is by subtract the number of correct answer in upper group from the number of correct answer in lower group. Then divide the difference proportion passing in the upper group and lower group by total number of students taking test in one group.

In analyzing the index of discrimination value, the writer used the similar steps with the steps used in index of difficulty. Such the writer applied the same way as steps that used in analyzing index of difficulty. First the writer calculated the student's score from the high score to the low score, and then the writer determined the class into two groups, upper and lower group based on the student's score calculated before. Next, the writer counted the index of difficulty used Heaton Formula as follow:

$$D = \frac{\text{Correct U} - \text{Correct L}}{n}$$

Then the last step the writer interpreted the result of scores conducted from analyzing the index of discrimination and categorized them using Heaton classification (stated in chapter III)..

After that, the writer categorized the result score of each items conducted from analyzing index of discrimination using Heaton's formula as stated in

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

previous chapter. Furthermore the writer applied Arikunto's classification to interpret index of discrimination value of each item test value, as follow:[6]

D : 00 up to 0,20 means it is poor

D : 0,20 up to 0,40 means it is satisfactory

D : 0,40 up to 0,70 means it is good

D : 0,70 up to 1,00 means it is excellent

However, the writer started analyzing from X1 class, X2, and then X3 class

1. Analyzing Index Discrimination on UAS Test for X1 Class

The result of analyzing index of discrimination of items test used in X1 as stated in the table 7 appendix 5 recorded that the big number of items tests in X1, it is 15 or 4.09 % of all items test, are poor since the most of items test have value discrimination between 0.06-0.24. They are number 1, 2, 15, 18, 20, 22, 23, 25, 27, 29, 30, 32, 35, 37, 40, 43, 47, 49, and 55. As Arikunto said that poor items dealing with index of discrimination must be revised because it cannot separating the good students from the bad.

The second huge numbers of items tests are categorized as satisfactory since they own discrimination value around 0.24-0.29. It is happen to 14 out of 55 items test, they are items number 3, 5, 7, 12, 13, 16,

[6] Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 221

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

27, 30, 36, 38, 41, 42, 45, and 46. Although they are satisfactory, they still need to be redesigned because they are doubtful to use in the other test. Besides, 6 items of all items test, they are items number 10, 21, 31, 33, 50, 51, produced value index discrimination 0.00. It means that these items did not work properly because they discriminate nothing. Thus, these items are must be deleted because they have no function at all or at least must be revised. Moreover, it became crucial since there are 10 items have negative sign value (-), means these items discriminate entirely in wrong way and they badly need to be retype because these items can mess the purpose of the test itself, it is to determine students who are more able and students who are less able.

The table 7 in appendix 5, complied from the result of the index discrimination analysis also recorded there are only 12 items that categorized good because they have discrimination value between 0.41-0.53, means these items are work properly to differentiate upper students from lower students and they could be utilized for the other test.

2. Analyzing Index Discrimination on UAS Test for X2 Class

The table 8 in appendix 6 reported that the items test used in this class are satisfactory since the big number of items test; it is 18 or 5, 8 % of all items test, they are number 5, 10, 16, 20, 24, 25, 26, 27, 30, 31, 33, 34, 36, 38, 39, 40, 42, 44, 48, and 51, have discrimination value between 0.22-

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

0.39. These items may to be refreshed in order they can functioned well. Because satisfactory items, as Heaton stated, are weak to discriminate upper level from lower level.

There are also 14 items are poor since they value between 0.06-0.17, they are items number 1, 2, 4, 7, 12, 13, 18, 19, 23, 29, 41, 45, 49, and 53. These items are must be revised because poor items cannot put the upper and lower student in the place they should be. Besides, there are 7 items number 8, 9, 11, 35, 46, and 52, did not function at all because their value is 0.00. These items are desirable to rewrite because they did not work at all and it will waste the time even money.

Moreover, there are 10 items that shown negative sign (-), means these items also must be revised since it entirely distinguished in wrong way. In the other hand, the table 8 in appendix 6 recorded that there are only 10 items that categorized as good items that can differentiate students in upper group from the lower group. They are items number 6, 14, 28, 54, and 55. Those items can be kept for the future test. But, however, it is still disappoint because the good items test related on index discrimination pointed only a few numbers.

3. Analyzing Index Discrimination on UAS Test for X3 Class

The results of index discrimination analysis due to X3 in the table 9 of appendix 7 recorded disappoint report since 18 or 5, 8 % of all items test

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

shown negative sign (-). It means that those items cannot be tested because they discriminate in false path especially items number 1, 2, 3, 5, 10, 12, 14, 15, 18, 23, 26, 27, 29, 35, 37, 44, 45, 46, and 49. They are crucial to be revised since they can manipulate the result of the test as expected, in the other word, these items can block the test to rich its purpose.

Moreover, 16 items test are poor. They are items number 9, 11, 16, 17, 19, 20, 25, 28, 32, 34, 38, 39, 47, 50, 53, and 54. These items have discrimination value between 0.05-0.17. As Arikunto stated that the items test which show index discrimination value between 0.00-0.20 are recognized as poor. It requires deeply thought to revise those tests because they cannot play proper work to discriminate upper students from lower students. Besides, there are 13 items test that are satisfactory because their facility value between 0.23-0.35. They are number 6, 7, 8, 13, 30, 31, 33, 36, 40, 41, 48, 51, and 52. Although they are satisfactory, but they still need to be refreshed in order they can work well. Because satisfactory items test are less good to facilitate item test to discriminate students. However, there are only 3 items test that have facility value 0.00. Means they do not function at all and extremely they must be deleted or at least revised. They are items number 4, 22, and 43.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

D. Analyzing the Effectiveness of Distracters.

The effectiveness of distracters is aimed to know whether the item test could work properly as expected or not. The result of analyzing distracters can be used either to revise the poor items and reference to design next other items test.

The analysis of destructors is done by comparing the number of students in upper group with students in lower group selected the false options given. In addition, the good distracters will manipulate more students whose belong to lower group than students belong to upper group. Thus, if there are more able students chosen the distracters, it means that the item does not function as expected in it must be revised.

The result from computing the effectiveness shown in the table 10-12 .in appendix 8-10, There are 5 columns in that table. First column contains the items number. Furthermore, English Final Test for Tenth Grade of SMAN 3 Sidoarjo has fifty five (55) items. The second column contains the alternative given in each items. In this case, there 5 options are available for each item, it is A, B, C, D, E. one alternative is the key answer and the rest are distracter. And it is added by 'omit' where in this case it is written by 'O'. Omit is used for both upper and lower students who chose nothing the incorrect alternatives.

In addition, the following explanations are used to describe the result of analyzing items distracters which is already written in that table.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1. Analyzing the Effectiveness of Distractors of UAS Test for X1 Class

The data for analyzing the effectiveness of distracters in the table 10 of appendix 8 reported that there are 66 non function distracters since non from both upper and lower of students chose those distracters. Besides, there are 27 distractors categorized as adequate, because they have same amount of voters from either better or poorer students. Moreover, there are 76 malfunction distractors since those items attracts more students in upper group than students in lower group, whereas good distractors should have been chosen by more bad students than by good students. [7] However, there 166 distractors are good since they have worked properly to mess less able student than able students. Thus, these distractors can be safety used in others items of future tests.

In addition, from the explanation above we can conclude that the distractors of Semester II English Final Test for Tenth Grade of SMAN 3 used in X1 are good since half distractors or 81. 50% of all distracters could function properly; it is attract more students in upper group than lower group of students. Then rest distractors are need to be revised or change because they cannot work properly and are not good to use for other tests. While those

[7] Burhan Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta, BPFE Yogyakarta, 2001). Page 144.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

adequate distracters are desirable may be used for the future tests without being revised.

2. Analyzing the Effectiveness of Distractors in X2 Class

Based on the result of analyzing the effectiveness of distractors of items test used for X2 shown in the table 11 of appendix 9, it can be seen that there are 130 distractors that good since they are chosen by more bad students than good students. It is relief because they could be kept and applied for others tests. There are also 127 non functioned distractors because no one from upper and lower student voted those alternatives. Thus, those distractors must be revised as Nurgiyantoro have said that if the distractors

Besides, there are 57 distractors are needed to be changed because they are malfunction. Those items are selected by more good students than bad students. However, good distractors should have attracted more students in lower group than students in upper group.

Furthermore, there are 16 distractors that adequate because the good students who chosen those distractors have the same amount with the lower students. According to Nurgiyantoro, those distractors are acceptable to use for future tests.

In conclusion, we simply say that the distractors of items test used for X2 are good since 133 out of 330 distractors performed efficiently to attract more

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

bad students than lower students. Hence, those distractors can be used for the next test. Nevertheless, this still disappoint since the non functioned distractors also shown big number, it is 104, or almost half distractors. Thus, the test makers or the teachers have to be more aware to revise those distracters.

3. Analyzing the Effectiveness of Distractors for X3 Class

Like in X2, the result of analyzing the effectiveness used for X3 as stated in the table 12 of appendix 10 recorded the similar report. There are 126 good distracters that can perform efficiently to attract more students in lower level. According to Nurgiyantoro, those items could be safety used for other tests.

There are also 107 items that has no function since they are not selected by both able and less able students. Thus, those distractors are not good to apply for other test and must be replaced.

Moreover, there are 78 distractors that worked in contrary as expected since they attracted the wrong candidates (i.e the better ones). Furthermore, there are 19 distractors that adequate because either upper or lower students have the same number in selecting those options.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

CHAPTER V

CONCLUSION AND SUGGESTION

In this chapter, the writer will present the conclusion and suggestion. In the conclusion, the writer will conclude the result conducted from the previous chapter, it is data analysis, including the result of analyzing content validity, index of difficulty, index of discrimination, and the effectiveness of distractors of Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo. While, in the suggestion, the writer would like to recommend matters that are need to do and not dealing with design good test to the several peoples stated in the first chapter, including teachers, students, and further researcher.

A. CONCLUSION

After discussing the result conducted from data analysis, we can conclude several matters

related to the agreement of content validity, index of difficulty, index of discrimination, and the effectiveness of distractors of Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo, as follow

1. From the result of analyzing content validity, it conducted that content validity of Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo has good content validity since the agreement of its content validity is 72%. It means that the items test covered almost all the material taught.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. The index of difficulty of Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo used for X1 is acceptable 27 or 13, 25% out of 55 items have facility value between 0.32-0.65 or categorized as good items. Thus, those items could safety used for future test without being to rewritten. While the rest items have to be revised because they are too easy and difficult. Moreover, since almost a half test, it is 24, shown difficulty value between 0.31-0.69, the items test used for X2 of SMAN 3 Sidoarjo are acceptable too. It because their facility value are recognized as good test. For items test that are easy and difficulty, they must be changed because they cannot work properly. Unlike X1 and X2, index of difficulty of UAS test used for X3 are easy since 21 out of 55 items test have difficulty value between 0.74-0.91. Hence, those items are not desirable to apply for other test and needed to be revised as well as the items that difficult.

3. Index of discrimination of UAS test used for X1 is poor since the biggest number of items, it is 15 out of 55, have discrimination value between 0.06-0.24. Those items badly need to be revised as well as the items that have satisfactory, non function, and even miss function index of discrimination. It is because all of those items cannot perform to distinguish better student from poorer student. While the rest items that have good index discrimination do not need to be revised and can be used for future test. Similarly, the index of discrimination of UAS test used for X2 are satisfactory since 18 out of 55

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

items test shown index discrimination value between 0.22-0.39. It indicates that those items need to be rewritten as well as those items are recognized poor, non function, and malfunction. While, 10 items that are categorized good are could be kept and applied for others tests. On the other hand, the index of discrimination in X3 are known malfunction since the biggest amount or items test, it is 18, have negative (-) sign. It means that those items cannot perform correctly because they discriminate the wrong candidate (better students). Thus, they badly need to be revised.

4. The result of analyzing the effectiveness of distractors of UAS test for X1, X2, and X3 class recorded the same report, that the items test are good since 166 or 81, 51% out of 330 distractors of items test for X1, 130 or 45% out of all distractors of items test for X2, and 126 or out of 330 distructors of items test for X3 are performed efficiently. It means that those distractors could safety used for future test, while the rest items that cannot work perfectly have to be revised.

B. SUGGESTION

After recognizing the result of this study related to content validity and items analysis of Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo, there are several matters that are seemed to be suggested.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

As stated in the first chapter, the suggestion due to the teacher, the students, and the further researcher, as follows:

For the teacher:

1. The teachers should try their own test out to know whether that tests are adequate before it is given to student as well as analyze it after it is tested. Hopefully, their test could perform correctly to measure the progress of their students as the function of good test.

2. Although the result of content validity of Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo is good, but teachers still should modify the items to be better items test since most of the items only cover one skill, it is reading, and 14 items were not taught in that class, moreover if the items test will be used once more in the future test.

3. Teachers should revise the items test that are poor, and difficult, in order there will no items that too easy and too difficult. Thus, teachers will know the progress of students correctly.

4. The teachers must aware to set the amount of items test for students. In the other word, we can say that amount of items test given must suitable with the time available for doing that test. If the items test are many, the time settled for work it must be added and vice versa. The amount of items test for Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo is not

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

suitable with the time available since the items are 55 and the time is only 90 minutes. It can influence student to rush the time left and much probably they will answer the items without carefully, and

5. Teachers should revise the items test that are poor, satisfactory, non function, mal function and even adequate, in order there will no items that discriminate in wrong way and could distinguish the better students from poorer students. Besides, teachers also should modify the distractors that are not good (non function, malfunction).

For students: Students should be able to recognize which tests those are good to do and which tests those are not good. Because bad test do not benefit them, in the other word, the bad test cannot measure their progress.

For further researcher: there must be several tests that are needed to be researched, in order to repair and fix all tests given and avoid from the designing and using of test that ignore of the criteria of good test (content validity, reliability, index of difficulty, index of discrimination, and the effectiveness of distractors).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BIBILIOGRAPHY

Arikunto, Suharsimi. 1996. *Prosedur Penelitian.* Jakarta: PT. Rineka Cipta.

Arikunto, Suharsimi. 1984. *Dasar-dasar Evaluasi Pendidikan.* Jakarta: Bina Aksara.

Bloom, Benjamin S.et all.. 1981. *Evaluation to Improve Learning.* USA

Brown, H. Douglas. 2001. *Teaching by Principles: An Interactive Approach to Language Pedagogy,* Second Edition. San Francisco State University: Longman, Inc

Cholbi, Laiyinatul. 2006. The Final English Test for the fourth grade students of SDN Pucang III Sidoarjo. Unpublished S-1 Thesis. Surabaya: UNESA (FBS)

Daryanto, Drs. H. 1999. *Evaluasi Pendidikan.* Jakarta: Rineka Cipta.

Depdiknas. 2004. *Standard Kompetensi Mata Pelajaran Bahasa Inggris SMP dan Madrasah Tsanawiyah.* Jakarta: Departemen Pendidikan Nasional

Ebel, Robert L. 1979. *Essential of Educational Measurement.* USA: Prentice-Hall INC New Jersey

Gary Borich, Tom Kubiszy. 2003. *Educational Testing and Measurement.* Singapore: John Wiley & Sons, INC.

Grounlund, Norman E. 1985. *Measurement and Evaluation in Teaching.* New York: collier Macmillan Publisher.

Hadi, Sutrisno. 1989. *Metodologi Research.* Yogyakarta: Andi offset.

Harris, David P. 1959. *Testing English as a second Language.* New York: Mc Craw-ill, Inc.

Hughes, Arthur. 2003. *Testing for Language Teacher.* Cambridge: University Press.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Heaton, J. B. 1988. *Writing English Language Test.* New York: Longman Group.

Lado, Robert. 1961. *Language Testing,* London: Longman Group.

Lampiran Peraturan Menteri Pendidikan Tentang Sistem Pendidikan Nasional No. 20 tahun 2003

Mardalis, Drs. 1995. *Metode Penelitian.* Jakarta: Bumi Aksara.

Nurgiyantoro, Burhan. 2001. *Penilaian dalam Pengajaran Bahasa dan Sastra.* Yogyakarta: BPFE Yogyakarta.

Purwanto, Drs. M. Ngalim. 1985. *Prinsip-prinsip dan Teknik evaluasi Pengajaran.* Bandung: Remadja Karya.

Peraturan Republik Indonesia Tentang Sistem Pendidikan Nasional No. 20 tahun 2003

Pratiktasari, Ria Dhewi. 2006. Semester II English Summative Test for The Eighth Year Student of SMP Negeri 1 Slahung Ponorogo. Unpublished S-1 Thesis. Surabaya: UNESA (FBS)

Rea-Dickins, Pauline& Kevin Germaine. 2008. *Evaluation.* New York: Oxford University Press.

Sumartana, Wayan Nurkanca, 1986. *Evaluasi Pendidikan.* Surabaya: Usaha Nasional.

Thoha, Drs. M. Chabib. 1991. *Teknik Evaluasi Pendidikan.* Jakarta: PT Raja Grafindo Persada.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id